

SCALEFREE HARDNESS OF AVERAGE-CASE EUCLIDEAN TSP APPROXIMATION

ALAN FRIEZE AND WESLEY PEGDEN

ABSTRACT. We show that if $P \neq NP$, then a wide class of TSP heuristics fail to approximate the length of the TSP to asymptotic optimality, even for random Euclidean instances. As an application, we show that when using a heuristic from this class, a natural class of branch-and-bound algorithms takes exponential time to find an optimal tour (again, even on a random point-set), regardless of the particular branching strategy or lower-bound algorithm used.

1. INTRODUCTION

In this manuscript, we prove that if $P \neq NP$, then *scalefree* heuristics cannot find asymptotically optimal approximate solutions even to random instances of the Euclidean TSP. Roughly speaking, *scalefree* heuristics are those which do not work especially hard at small scales. This has two important consequences.

First, it shows rigorously for the first time that several simple heuristics used for the TSP (Nearest Neighbor, Nearest-Insertion, *etc.*) cannot approximate the TSP to asymptotic optimality (even in average-case analysis), since these heuristics are all scalefree. In particular, our result can be seen as a defense of the intricacy of the celebrated polynomial-time approximation schemes of Arora [1] and Mitchell [15] for the Euclidean TSP, as we can show that the simpler algorithms cannot match their performance on sufficiently large instances.

The second consequence is a new view on the complexity of the Euclidean TSP versus other “actually easy” problems on Euclidean point-sets. Recall that for problems such as determining whether a Minimum Spanning Tree of cost ≤ 1 exists, or finding the shortest path between 2 points from a restricted set of allowable edges, the complexity status in the Euclidean case is unknown, as no algorithm is known to efficiently compare sums of radicals. In particular, it is conceivable that these problems are NP-hard (even if $P \neq NP$) as is the Euclidean TSP [16, 8]. Just as there are (sophisticated) efficient approximation algorithms for the Euclidean TSP, there are (trivial) efficient approximation algorithms for, say, the MST also: simply carry out Kruskal’s algorithm and calculate the length of the spanning tree to some suitable precision, to obtain a good approximation. Our result allows a rigorous distinction between these types of approximations: Kruskal’s algorithm is scalefree

Date: April 18, 2016.

Research supported in part by NSF grant DMS-1362785.

Research supported in part by NSF grant DMS-1363136.

in the sense of our paper, and we show that no algorithm with this property can well approximate the Euclidean TSP.

In particular, a surprising message from our result is that it is possible to connect Turing machine complexity to the practical difficulty of a problem for Euclidean point-sets, in spite of the unresolved state of the difficulty of comparing sums of radicals. In particular our result connects traditional worst-case deterministic analysis to *average-case*, *approximate* analysis of the Euclidean TSP, for a certain class of algorithms.

To motivate our definition of scalefree, we recall a simplification of the dissection algorithm of Karp, which succeeds at efficiently approximating the length of the shortest TSP tour in a random point set to asymptotic optimality.

- (1) Let $s(n) = \lfloor \frac{n^{1/d}}{\log^{1/d}(n)} \rfloor^d$ and divide the hypercube $[0, 1]^d$ into $s(n)$ congruent subcubes $Q_1, \dots, Q_{s(n)}$, letting $X_i = X \cap Q_i$ for each $i = 1, 2, \dots, s(n)$.
- (2) Using the dynamic programming algorithm of Held and Karp [9], find an optimal tour T_i for each set X_i .
- (3) Patch the tours T_i into a tour through all of X .

This algorithm runs in expected time $O(n^2 \log n)$ and finds an asymptotically optimal tour; i.e., the length of T is $(1 + o(1))$ times the optimal length, w.h.p. In some sense, its defining feature is that it works hard (running in exponential time in small sets of vertices) on a small scale, and is more careless on a large scale. Our definition of scalefree is intended to capture algorithms which do not exhibit this kind of behavior.

Roughly speaking, our definition of scalefree requires that the small-scale behavior of a heuristic can be efficiently simulated (usually, by simply running the heuristic on the local data). There are several caveats which must be kept in mind to avoid trivialities or impossibilities. For example, when simulating the local behavior of a heuristic, the simulation will have access to less precision than the heuristic does for the global set (since the precision must be a polynomial function of the data size); in particular, in some cases, correct simulation of the heuristic will be impossible at loss of precision, because the behavior of the heuristic is unstable at the local configuration. Our definition of scalefree thus only requires efficient simulation of local behavior to be possible *when the local behavior of the heuristic is stable*. The precise definition of scalefree, together with proofs that commonly used heuristics satisfy the definition, are given in Section 2. Our main theorem is as follows:

Theorem 1.1. *If $H(X)$ is a scalefree polynomial-time TSP heuristic and $P \neq NP$, then there is an $\varepsilon_H > 0$ so that w.h.p, $H(\tilde{X}_n)$ has length greater than $(1 + \varepsilon_H)$ times optimal, where \tilde{X}_n is a discretization of the random set $X_n \subseteq [0, 1]^d$ to any level of precision.*

For our purposes, a TSP heuristic H is an algorithm which takes an input set X , where $X = \{x_1, x_2, \dots, x_n\}$ is a list of points in $[0, 1]^d$. $H(X)$ returns such a tour, whose length we denote by $L_H(X)$. The coordinates of the points in X are

rounded to some precision of at most $\text{poly}(n)$ bits for input to H . For convenience, we consider $[0, 1]^d$ to be the d -dimensional *torus* instead of the hypercube; our proofs could easily be adapted to the hypercube.

For simplicity of notation, we will let $t = n^{\frac{1}{d}}$ and work with $\mathcal{Y}_n \subseteq [0, t]^d$, defined as the rescaling $t \cdot \mathcal{X}_n$. Observe that with this choice of rescaling, a typical vertex in \mathcal{Y}_n is at distance $\Theta(1)$ away from its nearest neighbor.

We will also consider the effectiveness of a scale-free heuristic in the context of a Breadth-First Branch and Bound algorithm. Defined precisely in Section 4, this is a branch-and-bound algorithm which branches as a binary tree, which is explored in a breadth-first manner. Branch-and-bound algorithms used in practice do typically satisfy this definition, as they make use of an LP-based lower bound on the TSP, and branch on the binary $\{0, 1\}$ -possibilities for fractional variables.

Theorem 1.2. *Any Breadth-First Branch and Bound algorithm that employs a scale-free heuristic to generate upper bounds w.h.p. requires $e^{\Omega(n)}$ time to complete.*

We close this section by noting that our proof of Theorem 1.1 provides a recipe to attempt to eliminate the $P \neq NP$ assumption for any specific, fixed scalefree heuristic H as follows. Let \mathcal{L} be any decision problem in NP.

Theorem 1.3. *For any scalefree polynomial-time TSP heuristic $H(X)$, there is a polynomial-time algorithm A_H such that for any $L \in \mathcal{L}$, we have:*

- (a) *If $L \notin \mathcal{L}$, then $A_H(L)$ returns false.*
- (b) *If $L \in \mathcal{L}$, then $A_H(L)$ returns either true, or exhibits an $\varepsilon_H > 0$ and a proof that $H(\tilde{\mathcal{X}}_n)$ has length greater than $(1 + \varepsilon_H)$ times optimal w.h.p.*

2. SCALEFREEDNESS OF HEURISTICS

Our notion of scale-freeness captures a common property of many simple TSP heuristics: namely, the small-scale behavior of the algorithm can be efficiently simulated (because it is essentially governed by the same rules as the large-scale behavior). To give a precise definition of scalefree, we let $B(p, r)$ denote a ball of radius r about p . For sets of points A, B , we write $A \approx_\varepsilon B$ to indicate that there is a bijection f from A to B such that for all $a \in A$, $\text{dist}(a, f(a)) < \varepsilon$.

We begin with the following notion:

Definition 2.1. Given $X \subseteq [0, t]^d$ and $Y = \{y_1, y_2, \dots, y_K\} \subseteq \mathbb{R}^d$, we say X is (R, ε) -protected by Y at p if

$$X \cap (B(p, R) \setminus B(p, 1)) = Y' \approx_\varepsilon (Y + p) \quad \text{and} \quad Y' \subseteq B(p, \sqrt{R}).$$

Our definition of scalefree is roughly that for local regions which are (R, ε) -protected by a suitable gadget Y (which can depend on the heuristic), the local behavior of the algorithm can be simulated in polynomial time, given only the local data. However, the precision of the local data will have to be rounded off (to be polynomial in the

number of points in the local data) and this rounding, in general, would make it impossible to predict the behavior of even simple heuristics such as Nearest-Neighbor, since they can be arbitrarily sensitive to the precise placement of points. It turns out for the proof of Theorem 1.1 that scalefreeness need only insist that the local behavior of a heuristic can be efficiently simulated *when the behavior of the heuristic is insensitive to perturbations in the points*. Thus, we define the ε -perturbation p_ε of a point $p \in [0, t]^d$ to be the random variable which is a point drawn uniformly randomly from the ball of radius ε about p , and, given a (finite) set $S \subseteq [0, t]^d$, define the random variable S_ε to be a set consisting of an independent ε -perturbation of each point in S .

Given a point $p \in [0, t]^d$, an ε -rounding of p is a point within distance ε of p ; the ε -rounding is *efficient* if it uses at most $\lceil \log(\frac{t}{\varepsilon}) \rceil$ digits for each coordinate. Note that efficient ε -roundings always exist for $\varepsilon > 0$.

For $X = \{x_1, x_2, \dots, x_n\}$, we let $X_k = \{x_i \mid i > k\}$. We begin with the definition of scalefreeness. (In the below, the role of \sqrt{R} is just that for large R , we have $1 \ll \sqrt{R} \ll R$.)

Definition 2.2. Let H be a polynomial-time TSP heuristic; it will be called *scale-free* if a certain implication holds. In particular, we require there to exist K, R, ε_1 , some *gadget* $Y = \{y_1, y_2, \dots, y_s\} \subseteq B(0, \sqrt{R}) \setminus B(0, 1)$ whose convex hull contains $B(p, 1)$ even after any ε -perturbation, and some polynomial-time algorithm $A = A_H$ which takes as input points in $[0, 2R]^d$, and which outputs a list of paths through the points. Then to be scalefree, H must satisfy the following implication:

If: We have:

- (a) $X \subseteq [0, t]^d$, $p \in [0, t]^d$, $0 < \varepsilon \leq \varepsilon_1$, $(X \cap B(p, 1)) = S \subseteq X_K$,
- (b) X is (R, ε) -protected by Y at p ,
- (c) the tour T_H found by H traverses $S \cup Y$ in a single path P_H ,
- (d) the vertices $x, y \in X \setminus P_H$ adjacent in T_H to the endpoints of P_H satisfy $\angle xpq, \angle ypq < \varepsilon_1$, where $q = p + (0, 1)$,

Then: When ε -roundings of the points of a congruent copy of S are given to A , we have EITHER that

- (i) H traverses S in a path, which belongs to the list produced produced by A when given as input ε -roundings of (a translation of) $X \cap B(p, R)$, OR
- (ii) If S, Y are replaced with the random perturbations S_δ, Y_δ for $\delta \geq Kd\varepsilon$, and the points in $X \setminus B(p, R)$ are moved arbitrarily inside $[0, t]^d \setminus B(p, R)$, then with probability at least .999 (with respect to S_δ, Y_δ), H will take a path other than P_H through the corresponding sets S_δ, Y_δ , OR
- (iii) the path taken by H through $S \cup Y$ can be shortened by ε .

Observe that the definition consists of an implication

$$(1) \quad (a) \wedge (b) \wedge (c) \wedge (d) \implies (i) \vee (ii) \vee (iii).$$

The definition could be strengthened (and simplified) by shortening the lists of requirements on the left-handside of the implication, and/or shortening the lists of allowable consequences on the right-handside of the implication. In particular, such

a stronger definition of scalefree would then immediately be subject to Theorem 1.1.

The complexity of the definition is thus just to ensure that it applies as widely as we would like. For example, we will see that the greedy heuristic satisfies the much simpler (and stronger) implication

$$(a) \implies (i) \vee (ii),$$

so this simpler notion of scalefreeness would suffice if we were only interested in the greedy heuristic.

The Nearest-Neighbor algorithm will satisfy the implication

$$(a) \wedge (b) \implies (i) \vee (ii),$$

which is more restrictive than our general notion of scalefreeness, but more permissive than the implication satisfied by the greedy algorithm.

We will use the full complexity (and weakness) of the implication (1) in the proofs that the Nearest-Insertion and Farthest-Insertion algorithms are scalefree.

More generally, the complexity of the definition can be understood from the fact that it links the difficulty of algorithms for Euclidean point-sets to the classical Turing machine complexity theory. The definition is required to allow loss of precision in local data so that a polynomial-time simulation is polynomial in the number of points of the local configuration, but compensates by only requiring simulation to be correct on relatively stable configurations (this is the role of allowing (ii)). It should not seem *a priori* clear that such a balancing act will succeed at allowing a result like Theorem 1.1.

The best motivation for Definition 2.2 comes perhaps from the proofs that many natural heuristics have this property. For simplicity, we consider just the case $d = 2$ in the proofs below; the proofs can be extended to higher dimensions with various degrees of effort.

2.1. Nearest Neighbor heuristic. The nearest-neighbor heuristic $\text{NN}(X)$ begins with the vertex x_1 , and, at each step, travels to the nearest vertex (up to calculation precision) which has not already been visited. To be polynomial time, distance calculations are cut-off at the same precision as the input, and the order of points is used to break ties. After visiting all vertices of X , the algorithm closes its path back to x_1 . Observe that despite the fact that distances are rounded to the same precision as the input, the heuristic will return a tour which is asymptotically equal in length to the length of the true nearest neighbor tour, provided that precision of the input is a sufficient function of $|X|$.

We will suppose that K is a sufficiently large integer, and let $Y = \{y_1, y_2, y_3\}$ be the vertices of an equilateral triangle with sidelengths 2 centered at the origin. Observe that taking $K \geq 1$ implies that we have that $x_1 \notin X \cap B(p, R)$; in particular, we have that the algorithm did not start from a vertex in $\hat{X} \cap B(p, R)$. Thus,

taking R sufficiently large and ε small ensures that the first visit of the algorithm to $\widehat{X} \cap B(p, R)$ is to a point of \widehat{Y}' , where Y' is the set given by the definition of (R, ε) -protected. Since R is large the algorithm will now exhaust the set $X \cap B(p, R)$ before leaving $B(p, R)$, and ignores points outside $B(p, R)$ until it does so. We will show that either case (i) or (ii) holds for S, Y ; case (iii) is unnecessary for the nearest neighbor heuristic.

In particular, aiming first for case (i), we aim to predict the paths it can take, given ε -roundings of $X \cap B(p, R)$. To do so, we run the nearest neighbor algorithm from each of the three possible initial entry points y_1, y_2, y_3 , to produce a sequence v_1, v_2, \dots . When, at step i , we encounter multiple candidates v_i which are closest to v_{i-1} up to precision ε , the algorithm branches, to attempt to catalog all possible nearest neighbor paths in the underlying data. We note that if x, y are both candidates for v_i at precision ε , then each of x, y would be closer to v_{i-1} after δ -perturbation of S with probability at least .49 (by making K large). Thus if the algorithm branches more than some fixed constant number of times, case (ii) would hold. As a consequence, we let our algorithm branch just some constant number of times, and are guaranteed that either case (i) or else (ii) holds.

2.2. Greedy algorithm. The greedy algorithm begins with T as an empty set of edges, and at each step, adds to T the shortest edge (up to precision) which does not create any cycle or vertex of degree 3. Ties can be broken using the index order of the endpoints of edges, lexicographically.

To prove scalefreeness, we take R, K sufficiently large, and Y as in the nearest-neighbor algorithm (Y could even be empty except that its convex hull must contain $B(p, 1)$), and ε small. Again, we will show that either case (i) or else (ii) hold. Using the greedy algorithm, the vertices in $X \cap B(p, 1) = X \cap B(p, R)$ will be covered by a path before any edges crossing the annulus $B(p, R) \setminus B(p, 1)$ are chosen by the greedy algorithm; in particular, the path taken by the greedy algorithm through $X \cap B(p, R)$ is independent of the placement of points outside of $B(p, R)$.

Moreover, observe that this path can be computed efficiently using the greedy algorithm on $X \cap B(p, 1)$ alone. As above, we branch when the choice is uncertain up to a distance of ε , and we find (for K sufficiently large) that when we branch at most some constant number of times to produce our list, we are either in case (i) or else in (ii).

2.3. Insertion Heuristics. Insertion heuristics build a tour by maintaining a cycle which is grown by inserting vertices one at a time. Several insertion heuristics exist. We will show that two are scalefree.

The *Nearest Insertion* algorithm begins, say, with T as the triangle on vertices x_1, x_i, x_j , where x_i is the nearest neighbor to x_1 and x_j is the vertex for which $d(x_1, x_j) + d(x_i, x_j)$ is minimized. At each step of the algorithm until T is a tour, the Nearest Insertion algorithm finds the vertex z in $X \setminus T$ which is closest to $V(T)$, finds the edge $\{x, y\} \in T$ for which $C = d(x, z) + d(y, z) - d(x, y)$ is minimized, and patches the vertex z in between x and y in the tour, at cost C .

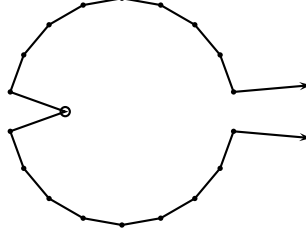


FIGURE 1. (Proving that the Nearest Insertion heuristic is scale-free.) The gadget Y (consisting of the 18 dots) prevents the optimal tour from entering directly to a vertex in $S \subseteq B(p, 1)$. The boundary of $B(p, 1)$ is drawn as the tiny circle.

To prove scalefreeness, we first observe that for any Y , S , we may assume that there is only ever one insertion of a vertex from $Y \cup S$ at an edge whose endpoints both lie outside of $Y \cup S$; if more than one such insertion occurs in the running of the heuristic, then the final tour will not intersect $Y \cup S$ at a path, violating assumption (c) the definition of scalefreeness.

We use the gadget Y shown in Figure 1. This consists of, say, 18 equally spaced points on the circle of radius \sqrt{R} with center $p + (0, \frac{\sqrt{R}}{2})$.

Now the path drawn in Figure 1 transits $S \cup Y$ optimally (assuming the route through S is optimal). Moreover, hypothesis (d) from the implication in Definition 2.2 ensures that the endpoints of this path are those used by the optimal tour. In particular, for sufficiently small $\varepsilon > 0$, we have either that (iii) holds, or else that the Heuristic traverses $S \cup Y$ as shown in Figure 1 (and transits S optimally).

Assuming this is the case, we aim to predict the precise path taken by the heuristic (in particular, in S). Our choice of Y ensures that the subtour built by the Nearest Neighbor heuristic will contain all of Y before it contains any vertex from S ; thus, we are guaranteed that the two edges leaving $B(p, R)$ are never used for insertions of points in $B(p, 1)$ (as there is always a cheaper insertion using closer edges). Since no edges with both endpoints outside of $B(p, R)$ are used for insertions after the first insertion, all insertions of points in S have both endpoints in $B(p, R) \cap X$. In particular, we can use the nearest insertion algorithm locally to determine the resulting path. We again branch when the next step is uncertain up to perturbation of the points by distance ε , and $(Kd\varepsilon)$ -stability (for sufficiently large K) again implies that when we branch at most some constant number of times, we achieve either case (i) or else case (ii).

We note that with small modifications, it is not hard to extend the scalefreeness proof to the *Farthest Insertion* heuristic, which inserts at minimum cost the farthest vertex from the vertex-set of the current subtour. The main problem is just that the heuristic will visit S before exhausting Y . However, it will only visit 1 vertex of S before exhausting Y , which means that we can (at polynomial cost) simply condition on the first vertex of S visited by the heuristic. \square

3. ASYMPTOTIC LENGTH OF SCALEFREE HEURISTICS

Our proof involves a multi-layered geometric construction, to force the appearance of certain special sets S satisfying hypotheses (a), (b), (c), (d) of the implication in definition of scalefreeness. We consider the layers one at a time.

3.1. Papadimitriou's set \mathcal{P} . Our proof begins with Papadimitriou's reduction [16] to the TSP path problem from the NP-complete set cover problem. This will form the basis for the sets S we wish to apply the definition of scalefreeness to.

Recall that an instance of the set cover problem is a family of subsets $A_i, i = 1, 2, \dots, m$ of $[n] = \{1, 2, \dots, n\}$; the decision problem is to determine whether there is a subfamily which covers $[n]$ and consists of pairwise disjoint sets.

In particular, Papadimitriou shows that for any instance (\mathcal{A}, n) of the set cover problem, there is a k (polynomial in the size of the set cover problem), and a set of k points $\mathcal{P} = \mathcal{P}(\mathcal{A}, n)$ in $[0, \sqrt{k}]^2$ with distinguished vertices p and q which can be produced in polynomial time, such that for some absolute constant ε_0 we have that for any approximator $f_{\varepsilon_0} : \mathcal{P} \rightarrow [0, \sqrt{k}]^2$ with $\text{dist}(f(x), x) < \varepsilon_0$ for all $x \in \mathcal{P}$ that

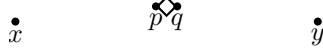
- P1 The shortest TSP path on \mathcal{P} begins at $f_{\varepsilon_0}(p)$ and ends at $f_{\varepsilon_0}(q)$.
- P2 There is a real number L such that the length of the shortest TSP path on $f_{\varepsilon_0}(\mathcal{P})$ is either less than L or greater than $L + \varepsilon_0$, according to whether the set cover instance problem should be answered Yes or No, respectively.

Papadimitriou's discussion does not reference an approximating function like f_{ε_0} ; the role of this function here is to capture the (im)precision which can be tolerated by the construction, which is discussed on page 241 of his paper (where one finds, for example, that we can take, e.g., $\varepsilon_0 = \frac{\sqrt{a^2+1}-a}{100(4a^2+2a)}$ for $a = 20$).

3.2. The set \mathcal{Q} . Papadimitriou's construction in [16] does not have the following property, but it is easy to ensure by simple modification of his construction (using "1-chains" to relocate the original p and q to suitable locations):

- P3 There is a rhombus R with vertices p, q so that all points in $f_{\varepsilon_0}(\mathcal{P} \setminus \{p, q\})$ lie inside R and at least ε_0 from the boundary of R .

When scaled to $[0, \sqrt{k}]^2$ as we have done here, the minimum TSP path length through Papadimitriou's set will always be less than $C_0 k$ for some absolute constant $C_0 > 1$ (indeed, this is true even for a worst-case placement of k points in $[0, \sqrt{k}]^2$ [6]). Thus, given the configuration $\mathcal{P} = \mathcal{P}(\mathcal{A}, n)$ with $k = k(\mathcal{A}, n)$ points and a small $\lambda > 0$, we rescale \mathcal{P} by a factor of say, $\frac{1}{\lambda} C_0 \sqrt{k}$, to produce a corresponding set of points $\bar{\mathcal{P}} \subseteq [0, \frac{\lambda}{C_0 \sqrt{k}}]^2$ which necessarily admits a TSP path of at most $\leq \lambda$; note that $\bar{\mathcal{P}}$ satisfies the same properties P1-P3 above, with ε_0 rescaled to $\frac{\lambda \varepsilon_0}{C_0 k}$.

FIGURE 2. \mathcal{Q} forces a path through a Papadimitriou path.

Finally, we modify this configuration (as indicated in Figure 2) by adding two points x, y to the set. With $\bar{\mathcal{P}}$ centered at the origin $(0, 0)$, we take $x = (-1, -\beta), y = (1, -\beta)$, where $\beta \gg \lambda/\sqrt{k} > 0$ is chosen sufficiently small so that p, q are the closest points on the rhombus R to x and y , respectively. Thus x and y are $\frac{\lambda\varepsilon_0}{C_0k}$ closer to p and q than to any other point in $\bar{\mathcal{P}}$. We call the resulting set $\mathcal{Q}(\mathcal{A}, n) \subseteq [-1, 1] \times [0, 1]$.

Essentially, the point set \mathcal{Q} ensures that any optimal path passing through it will transit the Papadimitriou set optimally, by ensuring the optimal paths will only enter/exit \mathcal{Q} where we expect.

Lemma 3.1. *Let (\mathcal{A}, n) be an instance of the set cover problem, and $\mathcal{Q} = \mathcal{Q}(\mathcal{A}, n)$ with $|\mathcal{Q}| = k$. There is a sufficiently large D_0 , such that **If** we have that*

- (i) $\mathcal{Q} \approx_\delta \mathcal{Q} \subseteq [0, t]^d$ for $\delta = \frac{\lambda\varepsilon_0}{10C_0k}$
- (ii) $w, z \in [0, t]^d$ with $\text{dist}(\{w, z\}, \mathcal{Q}) \geq D_0$,

then the shortest TSP path W from w to z in $\mathcal{Q} \cup \{w, z\}$ has the property that $W \setminus \{w, z\}$ is the shortest path covering \mathcal{Q} . In particular, $W \setminus \{w, z\}$ has endpoints x, y , and transits the approximate Papadimitriou set $P \approx_\delta \bar{\mathcal{P}}$ optimally.

Sets $\mathcal{Q} \approx_\delta \mathcal{Q}$ will serve as the sets S to which we apply the definition of scalefreeness. Very roughly speaking, we will eventually be aiming to contradict (i), since a polynomial-time algorithm to predict an optimal paths through \mathcal{Q} 's would seem to solve the set-cover problem in polynomial time.

Proof. Recall that \mathcal{Q} is constructed by adding two points x, y to the set $\bar{\mathcal{P}}$. By construction, the shortest path covering \mathcal{Q} has endpoints x, y , and is of length $< 2 + \lambda + 2\beta$. Moreover, it is apparent that any path covering \mathcal{Q} which does not have the endpoint pair $\{x, y\}$ has length at least 3. Finally, our choice of δ ensures that the accumulated error in path-lengths when comparing paths in \mathcal{Q} vs \mathcal{Q} is less than $2(k+1)\delta < \lambda$. Now we suppose that in the shortest path W , w is adjacent to α and z is adjacent to β , where $\{\alpha, \beta\} \neq \{x', y'\} \subseteq \mathcal{Q}$, where x', y' correspond to $x, y \in \mathcal{Q}$. Since W is shortest, we must have that

$$\text{dist}(w, \alpha) + \text{dist}(z, \beta) + 3 \leq \text{dist}(w, x') + \text{dist}(z, y') + 2 + 2\lambda + 2\beta,$$

and so

$$(2) \quad \text{dist}(w, \alpha) + \text{dist}(z, \beta) \leq \text{dist}(w, x') + \text{dist}(z, y') - 1 + 2\lambda + 2\beta.$$

Similarly, we have

$$(3) \quad \text{dist}(w, \alpha) + \text{dist}(z, \beta) \leq \text{dist}(w, y') + \text{dist}(z, x') - 1 + 2\lambda + 2\beta.$$

So we suppose now that (2) and (3) hold simultaneously. Moreover, let us assume without loss of generality that we have one of the following three cases:

Case 1: $\text{dist}(w, x') < \text{dist}(w, y')$ and $\text{dist}(z, y') < \text{dist}(z, x')$, or

Case 2: $\text{dist}(w, x') < \text{dist}(w, y')$ and $\min\{\angle wx'y', \angle wy'x'\}$ is at least $\gamma > 0$, for some γ depending on β .

Case 3: wx and zp is the shortest pair of independent edges joining $\{w, z\}$ to Q .

Before finishing the proof for each case, let's verify that for D_0 large, these cases do indeed cover all cases. If either $\min\{\angle wx'y', \angle wy'x'\}$ or $\min\{\angle zx'y', \angle zy'x'\}$ are at least γ , then we are already in Case 2, by appropriate choices of the labels w, z, x', y' from the available pairs. If on the other hand both $\min\{\angle wx'y', \angle wy'x'\}$ and $\min\{\angle zx'y', \angle zy'x'\}$ are at most γ , then either the angle from w to the center of Q to z is in $(\pi - \gamma, \pi + \gamma)$, and we are in Case 1 with the correct choice of which endpoints of Q are called x', y' , or the angle is less than γ , and we are in Case 3 with a suitable choice of labels.

Case 1: In this case, making D_0 large ensures that $\text{dist}(w, x') - \text{dist}(w, \alpha)$ and $\text{dist}(z, y') - \text{dist}(z, \beta)$ are bounded by a number arbitrarily close to 0, violating (2) (or (3), if we had flipped the roles of w and z). So this case cannot occur in simultaneously with (2) and (3).

Case 2: The second condition of this case implies that for any $\alpha \in Q$, we have that $\text{dist}(w, x') < \text{dist}(w, \alpha) + \lambda$ (for sufficiently large D_0), which allows us to modify (2) to the inequality

$$(4) \quad \text{dist}(z, y') - \text{dist}(z, \beta) \geq 1 - 3\lambda.$$

But z is at at least some fixed positive angle from the line through x, y (to which all points in Q are arbitrarily close). Thus, making D_0 large ensures that $\text{dist}(z, \alpha)$ varies by an arbitrarily small amount as we vary $\alpha \in Q$, contradicting (4).

Case 3: Recall that the shortest path on Q goes from x to p , then through $\bar{\mathcal{P}}$ optimally to q , and then ends at y . In particular, this path has endpoint pair $\{x, y\}$.

There are two candidates for the second-best pair of endpoints for admitting a short path: $\{x, p\}$, and $\{q, y\}$. For the pair x, p , the short path goes from x to y to q , through $\bar{\mathcal{P}}$ to p . The path for the pair y, q goes from y to x to p , through $\bar{\mathcal{P}}$ to q . Which of these choices gives rise to a shorter path depends on the precise rounding $Q \approx \mathcal{Q}$, but any path with a pair of endpoints other than $\{x, y\}$, $\{x, p\}$, or $\{q, y\}$ will be longer than both of these choices (here we are using P1). Note that both the shortest path (from x to y) as well as both of these choices transit $\bar{\mathcal{P}}$ optimally.

Thus, since wx and zp is the shortest pair of independent edges joining $\{w, z\}$ to Q , we must have either that the shortest path from w to z through Q uses these edges and takes the optimum path in Q from x to p , or else that it takes a path in Q which is shorter than this optimum path from x to p . But from above, we see that each such choice transits $\bar{\mathcal{P}}$ optimally. \square

3.3. The set \mathcal{M}_H . We use \mathcal{Q} to construct a larger set \mathcal{M}_H . \mathcal{M}_H consists of:

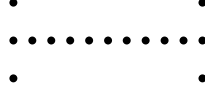


FIGURE 3. The set $\Pi(10)$ (rotated 90 degrees).

- (1) a copy of \mathcal{Q} , rescaled by a factor α to lie in $B(0, 1)$, and
- (2) the set $Y = Y_H \subseteq B(0, \sqrt{R})$ from the definition of scalefreeness.

The set \mathcal{M}_H ensures that hypothesis (b) of Definition 2.2 is satisfied at points centering approximations of \mathcal{M}_H .

The set Π_H . We now construct a certain set $\Pi = \Pi(k)$ and show that it constrains the optimal tour in a useful way. In particular, we let $\Pi(k)$ consist of the four points $\pi_1 = (0, 5)$, $\pi_2 = (0, 0)$, $\pi_3 = (1, 0)$, $\pi_4 = (1, 5)$ together with all the points $(\frac{1}{2}, \frac{5j}{k})$ for $0 \leq j \leq k$ (see Figure 3).

Lemma 3.2. *If $\text{dist}(\{x, y\}, \Pi(k))$ is sufficiently large and P is a shortest Hamilton path from x to y in $\Pi(k) \cup \{x, y\}$, then for at least one $i \in \{1, 2, 3, 4\}$ we have that neither neighbor v_i^1, v_i^2 of π_i on P is in $\{x, y\}$, and moreover that $\text{dist}(\pi_i, v_i^1)$ and $\text{dist}(\pi_i, v_i^2)$ become arbitrarily close to $\frac{1}{2}$ as k increases; in particular, the neighbors of π_i are nearly horizontal translates of π_i , lying on the line $x = \frac{1}{2}$.* \square

We now define Π_H as follows. We take four copies of the set \mathcal{M}_H , each rescaled to lie in small balls of radius $\varepsilon_\Pi > 0$. In $\Pi(k)$ (for $k = k(H)$ sufficiently large), we replace the four points π_i with these copies; those corresponding to π_3 and π_4 are reflected horizontally. (In particular, the resulting set Π_H still has horizontal reflection symmetry.)

Suppose we take this ε_Π suitably small, that $U \approx_{\delta_H} \Pi_H$ for sufficiently small $\delta_H > 0$, and that the optimal tour on $X \supseteq U$ transits U in a single path. Then each approximate copy of \mathcal{M}_H in U is transited in single path by the optimal tour (corresponding to hypothesis (c) from Definition 2.2) and moreover, Lemma 3.2 implies that for at least one of the four copies M_1, M_2, M_3, M_4 of \mathcal{M}_H , H either satisfies hypothesis (d) or else pays an additive error.

3.4. The set Π_H^3 . We let Π_H^3 denote 3 copies of Π_H centered at the vertices of an equilateral triangle of sidelength $2D_1$, say. This triple configuration ensures that the optimum tour will transit at least one of the copies of \mathcal{M}_H in a single pass. Indeed, Observations 2.9 and 2.10 from [7] now give the following:

Lemma 3.3. *Suppose that D_2 is a sufficiently large absolute constant, (\mathcal{A}, n) is an instance of the set cover problem, and $\Pi_H^3 = \Pi_H^3(\mathcal{A}, n)$. If $\Pi_H^3 \approx_{\delta_0} Z \subseteq X \subseteq [0, t]^d$ for $\delta_0 = \frac{\alpha \varepsilon_0}{10^4 C_0 D_0 k}$ and $\text{dist}(Z, X \setminus Z) \geq D_2$, then any TSP tour T on X can either be shortened in Π_H^3 by some additive constant or otherwise has the property that at least one of the (approximate) copies of Π_H in Z is traversed (optimally) by a path by T .* \square

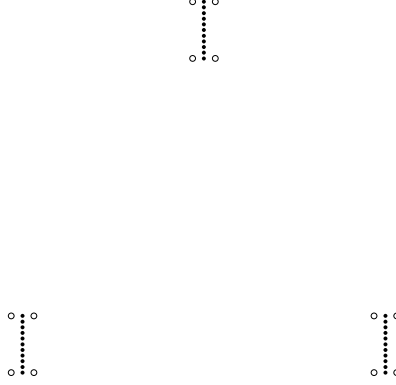


FIGURE 4. Π_H^3 , consisting of three copies of Π_H , forces a tour to transit one of 12 copies of \mathcal{M}_H (indicated here as the small circles) in a single pass, from a narrow prescribed angle.

We emphasize that the constants $\alpha \leq 1 \ll D_0 \ll D_1 \ll D_2$ are absolute, independent of (\mathcal{A}, n) .

At this point, we use the scalefree heuristic H to define the following polynomial time algorithm to solve a set-cover instance (\mathcal{A}, n) , which will be correct unless Theorem 1.1 holds for H . (Thus, $P \neq NP$ will imply Theorem 1.1.)

- (1) Compute $M = \mathcal{M}_H(\mathcal{A}, n)$ to precision $\frac{\delta_0}{4}$.
- (2) Produce a list of paths through M using the algorithm A from the definition of scalefreeness.
- (3) Let L' be the minimum length of a path covering one of the three Papadimitriou sets which is a subpath of one of the paths enumerated in step 2.
- (4) Return TRUE/FALSE according to whether L' lies below or above the threshold given by P2, respectively.

Proof of Theorem 1.1. We suppose that H is scalefree. $P \neq NP$ implies that there is some instance (\mathcal{A}_0, n_0) of the set cover problem for which the algorithm above returns an incorrect answer. Observe first that it cannot happen that the algorithm returns TRUE when the correct answer to the set-cover instance is FALSE: when (\mathcal{A}_0, n_0) is FALSE, property P2 implies that there can be no path through the Papadimitriou sets shorter than the threshold below which the algorithm above would return TRUE.

Thus we are to consider the case that the algorithm above returns FALSE even though the correct answer to (\mathcal{A}_0, n_0) is TRUE. In this case, no path enumerated by A transits M in such a way that a Papadimitriou set is traversed optimally.

We prove Theorem 1.1 by showing that this implies there exists an $\varepsilon_H > 0$ so that the length of the tour found by H through the random set $\mathcal{Y}_n \subseteq [0, t]^d$ is w.h.p at least $(1 + \varepsilon_H)$ times the length of the optimal tour T .

We begin with a definition. An (ε, R) -copy of S in $\mathcal{Y}_n \subseteq [0, t]^d$ is a set $S' \subseteq \mathcal{Y}_n$ such that $S' \approx_\varepsilon S$ and such that $\text{dist}(S', \mathcal{Y}_n \setminus S') \geq R$. We partition $[0, t]^d$ into cubes of sidelength $\approx 10dR$, say, color the cubes with 3^d colors so that no adjacent cubes get the same color, and call an (ε, R) copy of S in \mathcal{Y}_n R -aligned if it lies entirely in one of the subcubes in the first color class. We allow only at most one R -aligned copy per subcube, breaking ties by lexicographic order if necessary. Note that R -aligned copies are thus separated by distance $> R$. We have the following lemma (see Observation 3.1 from [7]).

Lemma 3.4. *Given any finite point set S , any $\varepsilon, \delta > 0$, and any R , we have that the number ζ_S of R -aligned (ε, R) -copies S' of S in a random set $X = \mathcal{Y}_n \subseteq [0, t]^d$ satisfies*

$$(5) \quad \zeta_S = C_{S,R,\varepsilon} n + o(n) \quad \text{w.h.p.}$$

for some constant $C_{S,R,\varepsilon} > 0$.

(Note that we require copies to be R -aligned just because it makes it easy to assert the $o(n)$ error term; we want more than just a w.h.p lower bound.)

Now we take ε_2 to be the minimum of $\frac{\delta_0}{8}$ and the parameter ε_1 from the definition of scalefreeness for the heuristic H , take $R = D_2$, take $S = \Pi_H^3 = \Pi_H^3(\mathcal{A}_0, n_0)$, and use Lemma 3.4 to find a linear number of $(\frac{\varepsilon_2}{(Kd+1)^{12}}, R + \varepsilon_2)$ copies Z of Π_H^3 which are $(R + \varepsilon_2)$ -aligned. In what follows, *aligned* copies will always refer to $(R + \varepsilon_2)$ -aligned copies.

We say that an aligned copy Z of Π_H^3 has the property Ξ_X if the tour T_H can be shortened within Z by δ_1 for some sufficiently small but fixed $\delta_1 > 0$. and we let ν_Ξ denote the number of aligned copies Z of Π_H^3 with property Ξ_X .

Claim: There exists $C > 0$ so that if H is scalefree, then $\nu_\Xi \geq Cn$ w.h.p.

Note that the claim immediately implies the theorem: in the rescaled torus $[0, t]^d$, the heuristic pays a total error of $\delta \cdot \nu_\Xi$, and rescaling by $t = n^{1/d}$, this gives Theorem 1.1. \square

Proof of the Claim. Each aligned copy Z consists of three copies $Z' \approx_\varepsilon \Pi_H(\mathcal{A}_0, n_0)$, and when Z fails to have property Ξ_X , Lemma 3.3 implies that at least one of the copies Z' is transited by T_H in a single path. Fixing a choice of such a copy Z' , Lemma 3.2 gives that at least one of the four copies M_i of \mathcal{M}_H in Z' (and so one of the twelve copies of \mathcal{M}_H in Z) satisfies hypotheses (c) and (d) of Definition 2.2. Therefore, on this copy M_i , we must have one of the three possible outcomes (i), (ii), or (iii) for the behavior of H in M_i . When Ξ_X fails for Z , we thus can fix a choice of such a copy M_i , which we call the *primal* copy M_i in Z , and we say that Z (without property Ξ_X) has property Λ_X if the primal M_i satisfies case (ii) of the

implication in Definition 2.2. We let ν_Λ be the number of aligned copies of Z with property Λ .

Earlier, we used Lemma 3.4 to find that there are w.h.p a linear number Cn of aligned $(\frac{\varepsilon_2}{(Kd+1)^{12}}, R + \varepsilon_2)$ copies Z of Π_H^3 ; these are thus also all (ε_2, R) copies. We assert now these copies must all have either property Λ_X or property Ξ_X (for sufficiently small δ_1).

Indeed, letting M_i be the primal copy of \mathcal{M}_H in Z , note that if case (iii) of the behavior of the scalefree heuristic H occurs at M' then we can indeed shorten T_H in M by an additive constant, and so have property Ξ_X . Also, if case (i) occurs at M_i , then our choice of \mathcal{A}_0 and Lemma 3.3 implies that we have property Ξ_X at Z . So otherwise case (ii) occurs at M_i , and Z has property Λ_X .

Thus far we have shown that

$$(6) \quad \nu_\Xi + \nu_\Lambda = (C + o(1))n \quad \text{w.h.p.}$$

for some $C > 0$. Next we will show that

$$(7) \quad \nu_\Lambda = (C' + o(1))n \quad \text{w.h.p.}$$

for some constant $C' \geq 0$, and that

$$(8) \quad \Pr(\nu_\Xi \geq C''n) > .0001.$$

The claim then follows from (6), (7), and (8), since (8) implies that $C - C' > 0$.

Equation (7) is immediate from the definition: since whether case (ii) is satisfied does not depend on the position of points in $X \setminus B(p, R)$, the number of R -aligned copies with property Λ_X is a sum of independent random variables, giving (7). Note that we do not assert that C' is positive.

We now aim to prove (8), by showing that if ν_Λ is large with high probability, then there is a positive probability that $\nu_\Xi \geq C''n$.

Now, abusing notation, we let $\varepsilon^{(1)} = \varepsilon_2$ and, more generally, define $\varepsilon^{(1)}, \varepsilon^{(2)}, \dots, \varepsilon^{(13)}$ by $\varepsilon^{(k)} := \frac{\varepsilon_2}{(Kd+1)^{k-1}}$. In particular, we have that $\varepsilon^{(k)} + Kd\varepsilon^{(k)} = \varepsilon^{(k-1)}$.

Now from X , we construct a sequence $X^{13}, X^{12}, \dots, X^1$ of random sets, as follows:

- X^{13} is the set X .
- Given X^k for $k > 0$, the set X^{k-1} is constructed by independently perturbing each point in X^k to a uniformly random point in the ball of radius $\varepsilon^{(k-1)} - \varepsilon^{(k)}$ about the point.

Note that this construction ensures that each X^k consists of $\delta^{(k)}$ -perturbations of the points in X , for $\delta^{(k)} = \varepsilon^{(k)} - \varepsilon^{(13)}$. Most importantly, observe that each X^k is identical in distribution to X . (This is where our use of the torus is convenient.)

Recall that Y_δ is obtained from a set Y by randomly perturbing each point in Y in a ball of radius δ . We have:

Observation 3.5. *Any event which occurs in the random set $X_{\delta(j)/2}$ with probability ρ occurs with probability $\geq \rho/10$ in the set X^j . More generally, any event which occurs in the random set $X_{\delta(j-k)}^k$ occurs with probability $\geq \rho/10$ in the set X^j ($13 \geq k > j \geq 1$).* \square

Recall that we found w.h.p linearly many $(\varepsilon^{(13)}, R + \varepsilon^{(1)})$ copies Z of Π_H^3 in X . Our choice of the $\varepsilon^{(j)}$'s is such that for each such Z , its corresponding perturbation Z^k in X^j is still an $(\varepsilon^{(j)}, R)$ copy of Π_H^3 , for each $13 \geq j \geq 1$. We aim to show that with positive probability, for at least one value of $k = 13, \dots, 1$, the corresponding $(\varepsilon^{(k)}, R)$ copy Z^k in $X^k = X_{\varepsilon^{(k)} - \varepsilon^{(13)}}$ has property Ξ_{X^k} , even conditionally on the locations of points in $X_{\varepsilon^{(k)}} \setminus B(p, R)$ where p is the center of Z .

Now our definition of the $\varepsilon^{(k)}$'s has been chosen so that, fixing k, ℓ , we have that $X_{\varepsilon^{(k)} - \varepsilon^{(13)}}$ is a random δ -perturbation of $X_{\varepsilon^\ell - \varepsilon^{(13)}}$ for $\delta \geq Kd\varepsilon^{(13)}$. Now fix some aligned $(\varepsilon^{(13)}, R + \varepsilon^{(1)})$ copy Z , and let \mathcal{B}_Z be the event that there exists any pair $13 \geq \ell > k \geq 1$ such that restriction of the path the Heuristic takes through the primal copy M in Z^ℓ is equivalent to the path taken through that same copy in Z^k . If Z has probability Λ then the conclusion (ii) from Definition 2.2 together with Observation 3.5 gives that the probability of \mathcal{B}_Z is at most

$$(9) \quad 10 \cdot \binom{13}{2} \cdot .001 < 1,$$

even conditioned on the positions of the points outside of Z .

On the other hand, let $M_{j_k}^k$ be the primal copy of \mathcal{M}_H in the copy Z^k in X^k , for each $k = 13, \dots, 1$. Since there are only 12 possible values for each j_k , we always have that $j_k = j_\ell$ for some $13 \geq \ell > k \geq 1$. From (9), we know that with positive probability, the Heuristic traverses either $M_{j_k}^k$ suboptimally in X^k , or traverses $M_{j_\ell}^\ell$ suboptimally in X^ℓ . In particular, we learn that if there were a linear number of aligned copies with property Λ_X , then for at least $C_0 n$ of the aligned $\varepsilon^{(13)}$ copies Z in X (for $C_0 > 0$), there is some k for which Z^k in X^k has property Ξ_{X^k} , with positive probability, even conditioned on the points outside of Z . Thus w.h.p, there is some $13 \geq k \geq 1$ such that $\nu_{\Xi}^k \geq C_0 n/13$, where ν_{Ξ}^k denotes the number of $\varepsilon^{(k)}$ copies of Π_H^3 in X^k with property Ξ_{X^k} . Now each ν_{Ξ}^k is identical in distribution to ν_{Ξ} , since the X^k 's are identical in distribution to X . In particular, we have that

$$\Pr(\nu_{\Xi} \geq C_0 n/13) \geq (1 - o(1))/13,$$

giving (8). \square

4. BRANCH AND BOUND

In our paper [7] we considered branch and bound algorithms for solving the Euclidean TSP. Branch-and-bound is a pruning process, which can be used to search for an optimum TSP tour. Branch-and-bound as we consider here depends on three choices:

- (1) A choice of heuristic to find (not always optimal) TSP tours;

- (2) A choice of lower bound for the TSP;
- (3) A branching strategy (giving a *branch-and-bound* tree).

For us a branch-and-bound tree is a rooted tree $T_{B\&B}$ where each vertex v is labelled by a 4-tuple $(b_v, \Omega_v, I_v, O_v)$. Here I_v, O_v are disjoint sets of edges and Ω_v is the set of tours T such that $T \supseteq I_v$ and $T \cap O_v = \emptyset$. The value b_v is some lower bound estimate of the minimum length of a tour in Ω_v e.g. the optimal value of the Held-Karp linear bound relaxation [5], [10], [11]. In addition there is an upper bound B , which is the length of the shortest currently known tour, found by some associated heuristic. This is updated from time to time as we discover better and better tours. If the root of the tree is denoted by x then we have $I_x = O_x = \emptyset$.

In [7] we allowed essentially any branching strategy. Given \mathcal{X}_n , we allowed any method to produce a tree satisfying the following:

- (a) When v is a child of u , $I_v \supseteq I_u$ and $O_v \supseteq O_u$.
- (b) If the children of u are v_1, \dots, v_k , then we have $\Omega_u = \bigcup_{i=1}^k \Omega_{v_i}$.
- (c) The leaves of the (unpruned) branch-and-bound tree satisfy $|\Omega_v| = 1$.

This process terminates when the set L of leaves of the pruned branch-and-bound tree satisfies $v \in L \implies b_v \geq B$; such a tree corresponds to a certificate that the best TSP tour found so far by our heuristic is indeed optimum. It is clear that if $v \in L$ then Ω_v does not contain any tours better than one we already know.

In [7] we concentrated on showing that even if we had access to the exact optimum i.e. letting $B = \lambda$, the minimum length of a tour, none of a selected set of natural lower bounds would be strong enough to make the branch and bound tree polynomial size. Note that this result does not depend on the branching process itself being efficient.

The aim of this section is to show that even if $b_v = \lambda$ then a certain branching strategy will fail. Unlike in [7], we cannot allow any branching strategy for our present result, as we might (though extreme computation in the branching process) find that we directly branch to a vertex w where I_w is exactly the set of edges of the shortest tour, giving $B = \lambda$, causing the algorithm to terminate, given that $b_v = B$ for all leaves of the tree.

It turns that to prove our result, we will need only a mild restriction on the branching strategies allowed.

- (1) A vertex v of our tree has two children w_+, w_- . Here $I_{w_+} = I_v \cup \{e\}$, $O_{w_+} = O_v$ and $I_{w_-} = I_v$, $O_{w_-} = O_v \cup \{e\}$ for some edge e .
- (2) The branch and bound tree is explored in a breadth first manner i.e. if the root is at level 0, we do not produce vertices of level $k+2$ until all vertices at level k have been pruned or branched on.

Note that this captures most branching strategies used in practice, which typically are using an LP-based lower bound on the length of the tour, and branching on the binary values possible for fractionally-valued variables in the linear program.

Theorem 1.2 will follow easily from the following claim: let H denote some scale-free heuristic and for a vertex v of $T_{B\&B}$ let $H(v)$ denote the length of the tour constructed by H , when it accounts for I_v, O_v . Let $\lambda(v)$ denote the length of the shortest tour in Ω_v .

Lemma 4.1. *There exist constants $\varepsilon_1, \varepsilon_2$ such that w.h.p. if vertex v is at depth at most $k_1 = \varepsilon_1 n$ then $H(v) \geq \lambda(v) + \varepsilon_2 n^{1/2}$.*

Proof. Let $\alpha_0 n$ be the minimum number of copies of Z with property Ξ_X promised by our analysis above, and let $\alpha_1 n^{1/2}$ be a lower bound on the penalty paid by our heuristic for each copy of Z . Then if $\varepsilon_1 = \alpha_0/2$ we have the lemma for $\varepsilon_2 = \alpha_0 \alpha_1/2$. \square

Here we have used the fact that at depth at most k_1 , w.h.p. there will be a linear number of copies of Z that are unaffected by I_v, O_v . These copies provide the necessary increases over the optimum.

It follows from Lemma 4.1 that for v at depth at most k_1 we have

$$H(v) \geq \lambda(v) + \varepsilon_2 n^{1/2} \geq B + \varepsilon_2 n^{1/2}.$$

This means that v is not a leaf. It follows that w.h.p. there will be at least $2^{k_1} = e^{\Omega(n)}$ leaves and Theorem 1.2 follows. \square

5. FURTHER WORK

From among the heuristics used in practice, the major omission from the present manuscript are the k -opt improvement heuristics, and their relatives (such as the Lin-Kernighan heuristic). Are they scalefree in our sense (or a related sense for which Theorem 1.1 holds)?

Theoretically, a natural question is whether our definition of a scalefree heuristic can be significantly simplified, while still allowing Theorem 1.1.

REFERENCES

- [1] S. Arora, Polynomial time approximation schemes for Euclidean Traveling Salesman and other geometric problems, *Journal of the Association for Computing Machinery* 45 (1998) 753-782.
- [2] D. Applegate, R. Bixby, V. Chvátal and W. Cook, *The Traveling Salesman Problem: A Computational Study*, Princeton University Press, 2007.
- [3] J. Beardwood, J. H. Halton and J. M. Hammersley, The shortest path through many points, *Mathematical Proceedings of the Cambridge Philosophical Society* 55 (1959) 299-327.
- [4] N. Christofides, The Traveling Salesman Problem, in *Combinatorial Optimization*, N. Christofides, P. Toth, A. Mingoza and C. Sandi, Eds., John Wiley and Sons, New York, 1971.

- [5] G.B. Dantzig, D.R. Fulkerson and S.M. Johnson, Solution of a large scale traveling salesman problem, *Operations Research* 2 (1954) 393-410.
- [6] L. Few, The shortest path and the shortest road through n points, *Mathematika* 2 (1955) 141-144.
- [7] A Frieze and W. Pegden, Separating subadditive Euclidean functionals, to appear in STOC 2016.
- [8] M. R. Garey, R. L. Graham, and D. S. Johnson, Some NP-complete geometric problems, in *Proceedings of the eighth annual ACM symposium on Theory of computing (STOC '76)*, ACM, New York, NY, USA, 10-22 (1976).
- [9] M. Held and R.M. Karp, A dynamic programming approach to sequencing problems, *Journal of the Society for Industrial and Applied Mathematics* 10 (1962) 196-210
- [10] M. Held and R.M. Karp, The Traveling Salesman Problem and Minimum Spanning Trees, *Operations Research* 18 (1970) 1138-1162.
- [11] M. Held and R.M. Karp, The Traveling Salesman Problem and Minimum Spanning Trees Part II, *Mathematical Programming* 1 (1971) 6-25.
- [12] D. Johnson, L. Mcgeoch and E. Rothberg, Asymptotic Experimental Analysis of the Held-Karp Traveling Salesman Bound, *Proceedings of the 7th Annual ACM Symposium on Discrete Algorithms* (1991) 341-350.
- [13] C. McDiarmid, On the Method of Bounded Differences, *Surveys in Combinatorics* 141 (1989) 148188.
- [14] D. Miller and J. Pekny, Exact Solution of Large Asymmetric Traveling Salesman Problems, in *Science* 251 754-761.
- [15] J. Mitchell, Guillotine Subdivisions Approximate Polygonal Subdivisions: A simple polynomial-time approximation scheme for geometric TSP, k-MST, and related problems, *SIAM Journal on Computing* 28 (1999) 1298-1309.
- [16] C. H. Papadimitriou, The Euclidean travelling salesman problem is NP-complete, in *Theoretical Computer Science* 4 (1977) 237-244
- [17] J. Michael Steele, Subadditive Euclidean functionals and nonlinear growth in geometric probability, *The Annals of Probability* 9 (1981) 365-376.
- [18] J. Michael Steele, Probability Theory and Combinatorial Optimization, SIAM CBMS series, 1996.
- [19] S. Steinerberger, New Bounds for the Traveling Salesman Constant, *Advances in Applied Probability* 47 (2013)
- [20] L. Tóth, Über einen geometrischen Satz, *Mathematische Zeitschrift* 46 (in German) (1940) 8385.
- [21] T. Volgenant and R. Jonker, A Branch and Bound Algorithm for the Symmetric Traveling Salesman Problem based on the 1-tree Relaxation, *European Journal of Operations Research* 9 (1982) 83-89.

DEPARTMENT OF MATHEMATICAL SCIENCES, CARNEGIE MELLON UNIVERSITY, PITTSBURGH, PA 15213, U.S.A.

E-mail address, Alan Frieze: `alan@random.math.cmu.edu`

E-mail address, Wesley Pegden: `wes@math.cmu.edu`